IMPROVING THE QUALITY OF MULTIMEDIA EXPERIENCE THROUGH SENSORY EFFECTS

Markus Waltl, Christian Timmerer, and Hermann Hellwagner

Multimedia Communication (MMC) Research Group, Institute of Information Technology (ITEC)
Klagenfurt University, Klagenfurt, Austria
E-mail: *firstname.lastname*@itec.uni-klu.ac.at

ABSTRACT

In previous and related work sensory effects are presented as a tool for increasing the user experience of multimedia presentations by stimulating also other senses than vision or audition. In this paper we primarily investigated the relationship of the Quality of Experience (QoE) due to various video bit-rates of multimedia contents annotated with sensory effects (e.g., wind, vibration, light). Therefore, we defined a subjective quality assessment methodology based on standardized methods. The paper describes the test environment, its setup, and conditions in detail. Furthermore, we experimented with a novel voting device that allows for continuous voting feedback during a sequence in addition to the overall quality voting at the end of each sequence. The results obtained from the subjective quality assessment are presented and discussed thoroughly. In anticipation of the results we can report an improvement of the quality of the multimedia experience thanks to the sensory effects.

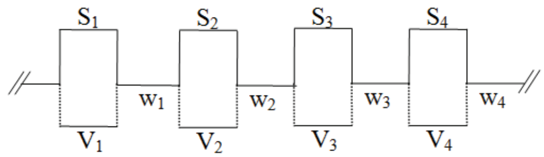
***Index Terms*—**Quality of Multimedia Experience, Sensory Effects, MPEG, Subjective Quality Assessment

1 INTRODUCTION

The past decade has witnessed a significant increase in the research efforts around the Quality of Experience (QoE), which is generally referred to as a human-centric paradigm for the Quality of Service (QoS) as perceived by the (end) user. Previous research has identified various dimensions in the area of QoE [1] while others work toward a theoretical framework [2]. One dimension recently gained momentum is 3D video [3] which aims for an increased immersive user experience [4].

In our research we cover yet another (new) dimension of the QoE, which we think, will go beyond current 3D technology that is entering the consumer market right now. In particular, we argue that the consumption of multimedia resources may stimulate also other senses than vision or audition, e.g., olfaction, mechanoreception, equilibrioception, or thermoception which shall lead to an enhanced, unique user experience. Therefore, the multimedia resources are enriched with additional metadata describing so-called sensory effects that are rendered on sensory devices like fans, vibration chairs, lamps, perfumer, etc. Note that sensory effects are not limited to installations, e.g., in home environments, there is already research to bring sensory effects to mobile devices [5]. In previous work we have implemented a test-bed [6] and performed first subjective tests using different sequences and genres [7]. The findings can be summarized as follows. Different ambient light settings can be achieved by deploying various (semi-) automatic color extraction algorithms within the rendering device at different complexities [6]. Sensory effects have been demonstrated as a vital tool for enhancing the user experience while the results are most promising for the genres documentary, action, sport, commercial, and news in decreasing order [7]. It shall be noted that for news the enhancement was insignificant compared to other genres.

The initial purpose of this paper was to investigate the relationship of the QoE of multimedia content annotated with and without sensory effects whereby the bit-rate of the pure multimedia content (i.e., without sensory effects) is equal to the bit-rate of the multimedia content including the sensory effect annotations. Due to the metadata overhead caused by the sensory effects we anticipated a significant difference in the objective and, thus, subjective quality of the multimedia content versions independent of their sensory effect annotations. However, it turned out that the metadata overhead was insignificant – thanks to the efficient design of the metadata format – resulting in a marginal delta peak signal-to-noise ratio (PSNR) which was below 0.1dB. Hence, we slightly modified our initial goal by investigating the influence of sensory effects on various bit-rates, and consequently qualities in terms of PSNR, of the actual video. In particular, we were interested in the subjective quality gap between video resources annotated with and without sensory effects at different bit-rates. That is, we investigated the QoE relationship assuming (1) with $br_{w/SE}$ denoting the bit-rate of the multimedia content including sensory effects and $br_{w/oSE}$ denoting the bit-rate of the same multimedia content without sensory effect annotations with the relationship of the bit-rates as shown in (2).



S_1, S_2, S_3, S_4	…	Test sequences (reference and modified)
V_1, V_2, V_3, V_4	…	Voting for corresponding sequence
w_1, w_2, w_3, w_4	…	Grey screen (5 – 10 seconds) with overall voting for corresponding sequence

Figure 1. Testing method.

$$QoE\left(br_{w/SE}\right) \geq QoE\left(br_{w/oSE}\right) \quad (1)$$

$$br_{w/SE} \leq br_{w/oSE} \quad (2)$$

Additionally, in [7] we had to adopt standardized test procedures to our needs. Therefore, we have started to work towards an alternative quality assessment method for sensory effects and additionally experienced with a novel voting device.

The remainder of the paper is organized as follows. Section 2 describes the test environment, its setup, and conditions including a novel voting device. The results are described in Section 3 and a concise discussion thereof is presented in Section 4. Section 5 concludes the paper and points out future work.

2 TEST ENVIRONMENT

2.1 Subjects and stimuli

In general, subjective tests are performed with around 10 up to 20 subjects in order to achieve acceptable results. For our tests we have invited 24 students (11 female and 13 male) between the age of 18 and 37 years. Two students took already part in our previous experiments [7] but overall they were neither familiar with the evaluation topic nor subjective assessments in general.

In contrast to our previous work [7] we have prepared a written introduction explaining the test procedure in detail including a description of the voting device (cf. Section 2.3) and the voting scale [8].

The test stimuli comprise prepared video sequences as described in Table 1 whereby “Babylon A.D.” depicts a trailer of an action movie and “Earth” shows a documentary. The former is characterized with a lot of shot/scene transitions and fast motion whereas the latter has much less shots/scenes and also less motion. The two sequences were chosen because they are the ones with the best results from our previous experiments [7]. For each sequence four versions with different bit-rates are prepared whereby only the video bit-rate is affected and the audio bit-rate remains constant for all versions within a sequence. Note that the versions identified as highest quality are used as references for calculating the average PSNR and therefore no PSNR values are provided. The quality of the videos can be summarized like the following: the lowest bit-rate had clearly visible artifacts, i.e., at explosion, fast movement, fades and contiguous block artifacts throughout the whole sequence. The medium quality had also visible block artifacts but these were not as strong as in the lower quality. Block artifacts in the high quality version of the sequences nearly disappeared and became visible only during fast motion (e.g., explosions). The highest quality had no block artifacts within the whole sequence. Additionally, each sequence has been annotated with sensory effects (wind and vibration; light effect are automatically extracted as described in [6]) according to [9] resulting in 16 different bit-streams to be evaluated. It is also noted that the duration of the sequences was longer (35s and 21s) than that usually used within subjective tests (~10s). The reason for this was to allow for having more sensory effects within a single test sequence and to accommodate requests from subjects in our previous tests which have been provided in a post-experiment questionnaire [7]. Furthermore, Table 1 presents the number of wind (W) and vibration (V) effects for each sequence. The light effects were calculated automatically every 0.1 second which approximately led to "video duration / 0.1" light effects.

Table 1. Video sequences and PSNR/bit-rate versions.

Sequence	Babylon A.D.		Earth	
Duration	35s		21s	
Resolution	1280 x 544		1280 x 720	
Motion	High		Low	
Nr. of Effects	W: 7; V: 9		W: 8; V: 1	
Bit-rates	**Kbit/s**	**PSNR**	**Kbit/s**	**PSNR**
Low Quality	2154	38.93	2204	38.11
Medium Quality	3112	41.27	3171	40.65
High Quality	4044	42.95	4116	42.27
Highest Quality	6315	N/A	6701	N/A

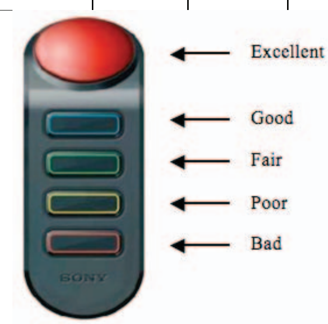


Figure 2. Voting device and mapping to voting scale.

2.2 Test method and experimental design

For accomplishing the subjective tests the following hardware and software was used:

- Dell Precision 360: Pentium 4 3.2 GHz with 1 GB RAM and NVidia Quadro NVS (64 MB)
- amBX Premium Kit (Fans, Vibration bar, Lights, Sound)
- 26’’ Monitor with a resolution of 1680x1050
- Voting device (cf. Section 2.3) and JoyToKey 3.7.4
- Windows XP SP3
- Sensory Effect Media Player (SEMP)

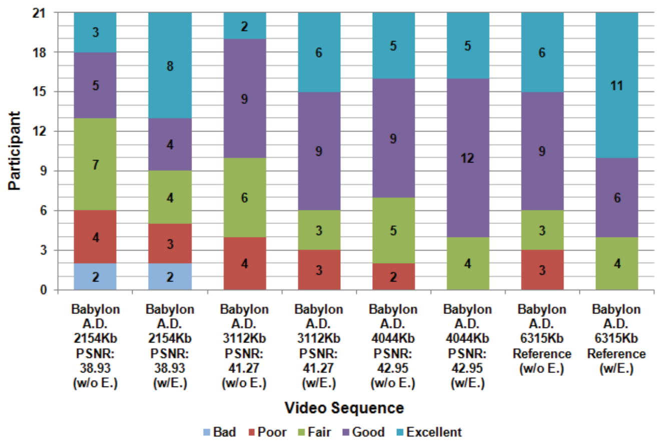


Figure 3. Evaluation results for Babylon A.D.

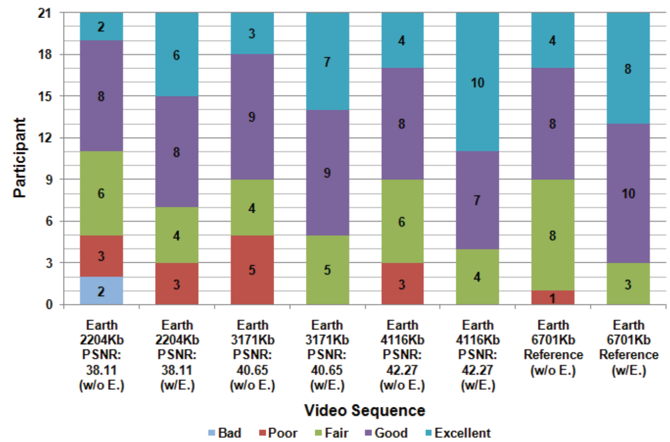


Figure 4. Evaluation results for Earth.

- amBX Software (i.e., amBX System 1.1.3.2 and Philips amBX 1.04.0003)

The test setup comprises a control station and the actual test computer. The test computer is equipped with the amBX premium kit comprising a wall washer light with controller unit, left & right 2.1 sound speaker lights and a sub woofer, a set of fans, and a wrist rumbler. The wall washer includes high power RGB LEDs with over 16 million additive RGB colors, instant response, and continuously variable intensity. The 2.1 sound system comes with 160 W music power, two 40 W satellites with light devices on top of it and a 80 W subwoofer operating in the frequency range of 35 Hz ~ 20 kHz. The two fans have variable speed control with up to 5,000 rpm. Finally, the wrist rumbler has two integrated motor drives with variable rotation speed. The control station is used to start the scripts for the sequences. Furthermore, it is used to restart SEMP if any issues arise.

The actual test was divided into three parts with a total duration of around 25 minutes per subject. The *first part* comprised the introduction where each subject had to read the document explaining the test procedure.

The actual subjective assessment was conducted within the *second part*. Therefore, the subjects sat (in a comfortable seat) in a distance of around three times the height of the monitor and the two different sequences with and without sensory effects (i.e., four in total) have been presented to the subjects in randomized order. As each sequence had four different bit-rate versions – i.e., also randomized order within a sequence – the subjects had to evaluate 16 videos in total. The test method for one sequence is shown in Figure 1 and the reference sequence, i.e., the one with the highest quality, is hidden. Thus, we have adopted the Absolute Category Rating with Hidden Reference (ACR-HR) method for our evaluation [10]. The major difference in our test method was that subjects were able to vote during the sequences thanks to our novel voting device (cf. Section 2.3). Between every sequence there was a five seconds break to allow the participant to give the overall quality rating. The actual scale was a five-point discrete scale from excellent to bad as defined in [10]. Additionally, we presented the subjects a small scale on the bottom-right corner of the video. This scale gave a voting feedback providing the last given vote from the subjects.

After all sequences were displayed and evaluated the subjects had to answer the post-experiment questions within *part three* of the test. For this part the participant had no time limit and could ask questions about the questionnaire. The following questions were asked during the post-experiment part:

Q1. How easy or difficult was it to determine the impairment of the video?
Q2. Would you have liked less or more time to hear-see the sequence with sensory effects?
Q3. Was the presented voting feedback disturbing?
Q4. Did you direct your attention to any specific sensory effect when determining the quality of the experience?
Q5. Where you ever mentally overloaded during any part of the experiment?
Q6. Have you ever participated in an experiment similar to this one?
Q7. Any other comments about what you liked or did not like, or things that should be changed during the course of this experiment?

Finally, the overall test setup was inspired by and partially based on [11].

2.3 Voting device

As already introduced in the previous section we adopted a novel voting device, i.e., a Buzz! controller known from various gaming consoles. We mapped the individual buttons to the given five-point voting scale as depicted in Figure 2. For the actual mapping to a key input device and, consequently, also for the logging functionality, we used the JoyToKey[1] software library.

[1] http://www.electracode.com/4/joy2key/JoyToKey English Version.htm

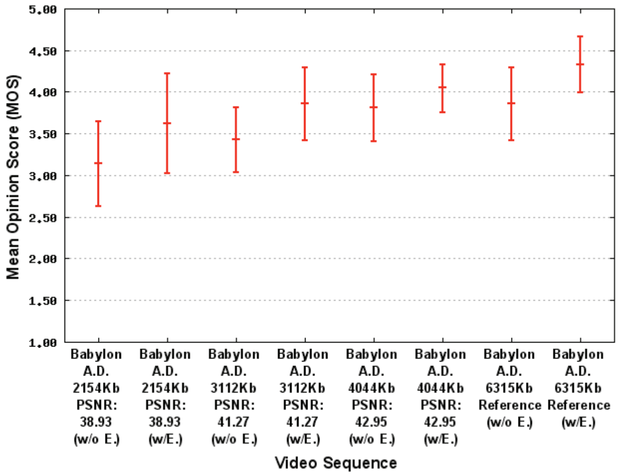


Figure 5. MOS and confidence interval for Babylon A.D.

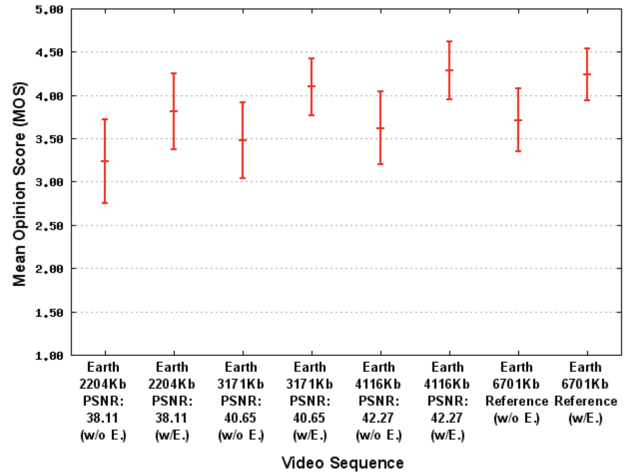


Figure 6. MOS and confidence interval for Earth.

The Buzz! controller gives the subjects enough freedom to vote during the sequences (i.e., for continues feedback) as well as after the sequences (i.e., for the overall quality) without modifying their position within the comfortable seat. Furthermore, the device allows the participants to easily adjust their vote by simply pressing a button instead of determining a specific position on a slider. The results for both voting possibilities are presented and discussed in the subsequent sections.

3 RESULTS

This section comprises the test results from our 24 subjects. Please note that we have identified three outliers which have been eliminated from the evaluation according to [12].

3.1 Overall evaluation results

Figure 3 and Figure 4 depict the voting results for the overall quality given by the subjects for the test sequences. Furthermore, it shows the rating separated by the PSNR/bit-rates and by sensory effects. A sequence annotated with sensory effects is indicated with *w/E.* and without sensory effects with *w/o E.* The results show that the quality of a video with sensory effects is rated higher than without sensory effects.

The results clearly indicate that the votings for *excellent* and *good* are higher for sequences annotated with sensory effects than for sequences without sensory effects.

As the voting device has not five equally sized and colored buttons it has to be mentioned that there was no visible influence on the voting process by the device itself (cf. Section 3.4).

3.2 Mean Opinion Score results

Figure 5 and Figure 6 show the mean opinion score (MOS) and confidence interval (95%) for both sequences with their variation in PSNR/bit-rate and sensory effects (with or without sensory effects). Interestingly, the sequences with sensory effects have always a higher MOS than their counterparts without sensory effects and almost steadily increase for higher PSNR/bit-rates.

3.3 Continuous voting results

As stated before, we asked the subjects to rate the quality of the video sequence during the playback. For the evaluation we split up the video sequences into a set of scenes. The scenes were selected by determining the length of shots and the motion within them. In particular, we did not select each shot as a scene if the shot had only a length below a certain threshold. Furthermore, using short scenes would lead to inaccurate results because the subjects were mentally overloaded and not able to vote during this short time period (cf. Section 3.4). For each scene we calculated the average MOS. Note that some scenes are longer than others resulting in more votes for the longer scenes. For example, the lowest number of votes is five for the interval six to eight for 4116 Kbit/s with sensory effects for the sequence *Earth* which is at the lower boundary of the statistical relevance.

Due to space constraints only two results are presented which have been selected as representative examples (cf. Figure 7 and Figure 8). As one can see, sometimes the sequence without sensory effects has a higher MOS value than its counterpart with sensory effects. This can be explained that some subjects were mentally overloaded during the test, especially in case the sequence had a lot of scene/shot transitions (cf. Section 3.4).

3.4 Post-experiment questionnaire results

The results of the post-experiment are as follows. For determining the quality (i.e., Q1), 9.52% stated that it was *very easy* to determine the quality of the video, 38.1% of the participants declared that it was still *easy*, 28.57% indicated that it was *difficult*, and 4.76% pointed out that it was *very difficult*. Concerning the timing (i.e., Q2), the participants

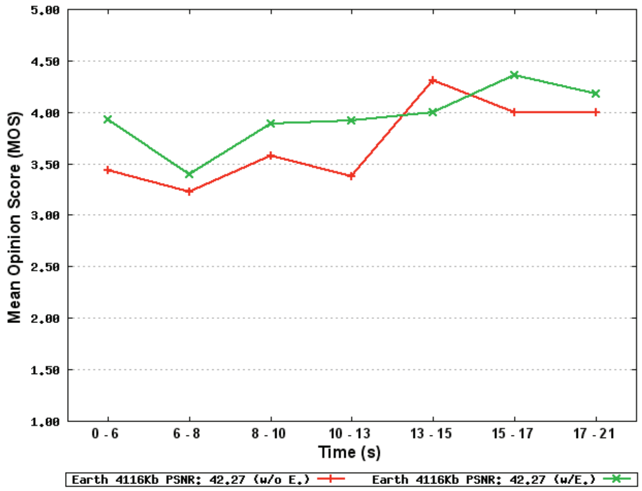


Figure 7. Continuous MOS for Earth at 4 Mbit/s.

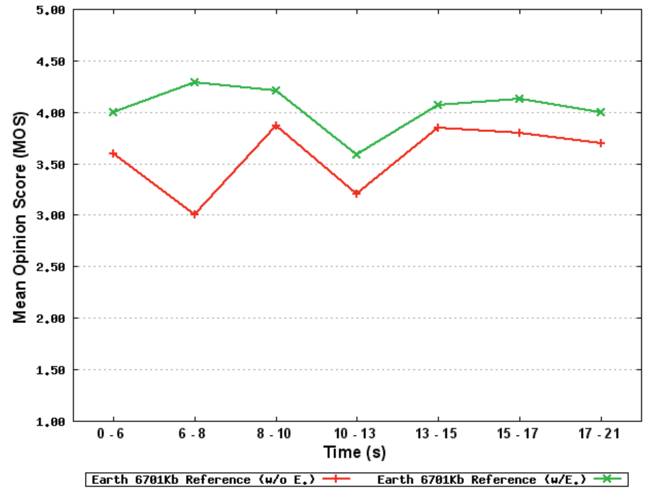


Figure 8. Continuous MOS for Earth at 6 Mbit/s.

indicated that they wanted to have *much* (4.76%) or *little more* (47.62%) time to hear/see sequences with sensory effects. Only a small number of people (9.52%) stated that they wanted to have *less time* to hear/see such sequences. Note that no one wanted to have *much less time* to hear/see sequences with sensory effects.

The presented voting feedback (i.e., Q3) in the bottom-right corner of the movie was found disturbing by 14.29% of the subjects. In contrast to the first subjective test the subjects equally directed their attention to wind, vibration and light (i.e., Q4). Furthermore, the subjects liked the buzzer for the voting because it did not interfere too much during watching the sequences. One problem for some subjects was the voting during the sequence of *Babylon A.D.* The reason for this was the high motion and the short scenes within the sequence. This resulted in stress for the subjects during the user study (i.e., Q5). For Q6 and Q7 no feedback worth mentioning here has been provided.

4 DISCUSSION

In general, the results reveal that sensory information in form of sensory effects associated to multimedia content is an appropriate tool for increasing the user experience by stimulating also other senses than vision and audition. We did also a cross-check of the results with our previous tests [7] which confirms the results reported in this paper. In particular, it seems that sensory effects are more appreciated for documentaries than for action movies (see higher MOS values as reported in Sections 3.1 and 3.2).

Another observation is that for each sequence and bit-rate version the MOS value is higher if annotated with sensory effects than without sensory effects (cf. Section 3.2). Additionally, MOS values increase with higher PSNR/bit-rate but this sounds quite obvious and was also expected. Therefore, we have compared the MOS values versus the PSNR/bit-rates for the two video sequences as depicted in Figure 9 and Figure 10 respectively. We calculated the average difference between the two curves using the Bjontegaard Delta (BD) method [13] with the following results. For the sequence *Babylon A.D.* the rating with sensory effects is around 0.4 MOS points higher than without sensory effects and for *Earth* the rating for the sequence enriched with sensory effects is with 0.6 MOS points higher than without sensory effects. One could interpret that the subjective quality for documentaries is higher than for the action genre. However, the difference between the two sequences lies also in the motion and number of scene/shot transitions whereby the action trailer (Babylon A.D.) has higher motion and higher number of scene/shot transitions than the documentary (Earth).

High motion in combination with many sensory effects could lead to mentally overloaded subjects during the assessment. The continuous voting results (cf. Section 3.3) in combination with the post-experiment questionnaire confirm this observation. We will address this issue in our future work (cf. Section 5).

Finally, we conclude that video resources annotated with sensory effects show an improvement of about 0.5 MOS on average compared to video resources without sensory effects. Furthermore, for the action trailer (Babylon A.D.) one can observe (cf. Figure 5) that the MOS of the lower bit-rate with sensory effects is always higher than the MOS of the next higher bit-rate without sensory effects although the difference is not that significant. Interestingly, for the documentary (Earth) the MOS of the lowest bit-rate with sensory effect is always higher than the MOS of all higher bit-rates without sensory effects (cf. Figure 6). Thus, one could conclude that the relation as shown in (1) and (2) is always true for the sequences and bit-rate variations used in the experiments discussed above. However, it is also clear that more tests are required to provide confident results for a wide range of sequences/genres which we will address as part of our future work.

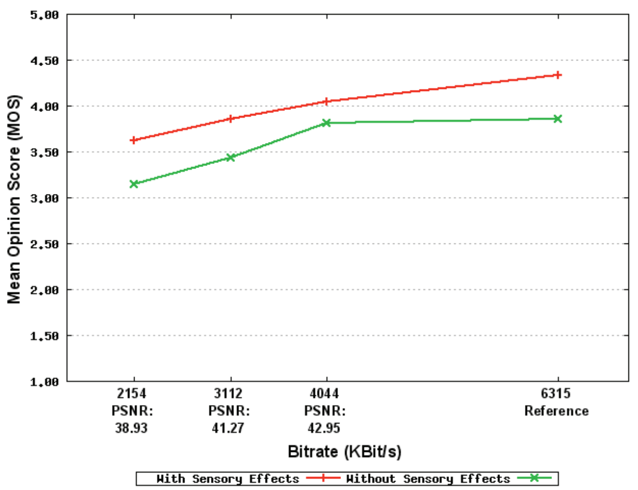


Figure 9. MOS vs. PSNR/bit-rate for Babylon A.D.

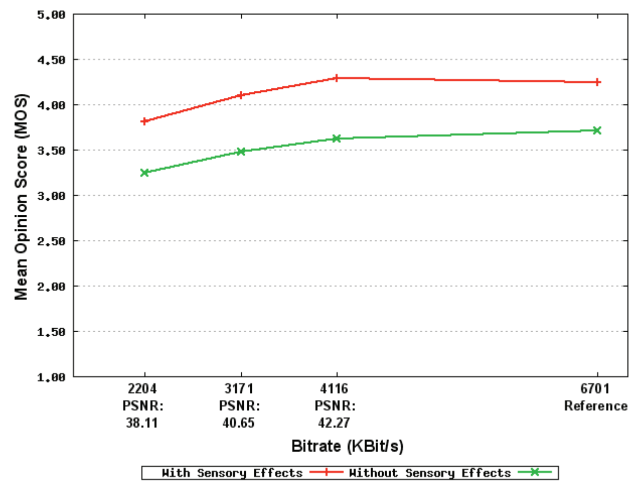


Figure 10. MOS vs. PSNR/bit-rate for Earth.

5 CONCLUSIONS AND FUTURE WORK

In this paper we have presented our results of a formal subjective quality assessment in the area of sensory information. Due to the nature of the sensory information we adopted and slightly modified one of the traditional assessment methods and we experimented with a novel voting device. The results are presented in detail and discussed within the previous sections. In particular, the results reveal that there is an improvement of the quality of the multimedia experience thanks to the sensory effects. However, it is still too early to quantify the QoE improvement in general due to the lack of standardized test sequences and test methods for the QoE assessments of multimedia content annotated with sensory effects.

In the future we will improve the test method by adopting a continuous quality scale instead of discrete scale. Furthermore, we will continue experiments with alternative voting devices and we will assess the degree of QoE improvement due to the influence of a single source of additional information (e.g., testing only light or wind). Next to that we will work towards a database of test sequences for single and combined effects at different bit-rates and perform intensive tests also across all ages with the ultimate goal to establish a generalized utility model for sensory information.

6 ACKNOWLEDGMENT

This work was supported in part by the EC in the context of the NoE INTERMEDIA (NoE 038419), the P2P-Next project (FP7-ICT-216217), and the ALICANTE project (FP7-ICT-248652).

7 REFERENCES

[1] F. Pereira, “A triple user characterization model for video adaptation and quality of experience evaluation,” *7th Workshop on Multimedia Signal Processing (MMSP) 2005*, Shanghai, China, Oct. 2005.

[2] W. Wu, A. Arefin, R. Rivas, K. Nahrstedt, R. Sheppard, Z. Yang, “Quality of experience in distributed interactive multimedia environments: toward a theoretical framework”, *ACM Multimedia 2009*, Beijing, China, Oct. 2009.

[3] A. Smolic, K. Mueller, P. Merkle, P. Kauff, T. Wiegand, “An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution”, *Picture Coding Symposium (PCS’09)*, Chicago, USA, May 2009.

[4] M. Barkowsky, R. Cousseau, P. Le Callet, “Influence of depth rendering on the quality of experience for an autostereoscopic display”, *1st Int’l Workshop on Quality of Multimedia Experience (QoMEX’09)*, San Diego, Jul. 2009.

[5] A. Chang, C. O'Sullivan, “Audio-haptic feedback in mobile phones”, ACM *Conference on Human Factors in Computing Systems (CHI'05), CHI '05 Extended Abstracts on Human Factors in Computing Systems*, Portland, OR, USA, pp.1264-1267.

[6] M. Waltl, C. Timmerer, H. Hellwagner, "A Test-Bed for Quality of Multimedia Experience Evaluation of Sensory Effects", *1st Int’l Workshop on Quality of Multimedia Experience (QoMEX’09)*, San Diego, USA, Jul. 2009.

[7] M. Waltl, C. Timmerer, H. Hellwagner, "Increasing the User Experience of Multimedia Presentations with Sensory Effects", *11th Int’l Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS’10)*, Desenzano del Garda, Italy, Apr. 2010.

[8] Introductory text for subjective quality assessment methods in the context of Sensory Information, http://bit.ly/bGibmG. (last accessed: Apr. 2010)

[9] ISO/IEC 23005-3 FCD, “Information technology – Media context and control – Sensory information”, *ISO/IEC JTC 1/SC 29/WG 11/N10987*, Xi’an, China, Oct. 2009.

[10] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications", Apr. 2008.

[11] R. L. Storms, M. J. Zyda, "Interactions in Perceived Quality of Auditory-Visual Displays", *Presence: Teleoperators and Virtual Environments*, vol. 9, no. 6, Dec. 2000, pp.557-580.

[12] ITU-R Rec. BT.500-11, "Methodology for the subjective assessment of the quality of television pictures", 2002.

[13] G. Bjontegaard, "Calculation of average PSNR differences between RD curves", *ITU-T VCEG Meeting, document VCEG-M33*, Austin, USA, 2001.